Third-party punishment: altruistic and anti-social behaviours in in-group and out-group settings

Journal of Cognitive Psychology

Third-party punishment: altruistic and anti-social behaviours in in-group and out-group settings

Volume 28, Issue 4, 18 May 2016, Pages 486-495

Rabellino, D., Morese, R. , Ciaramidaro, A., Bara, B.G., Bosco, F.M.

Title (96/150): Third-Party Punishment: Altruistic and anti-social behaviors in in-group and out-group settings

Authors: Daniela Rabellino[1,2], Rosalba Morese[1], Angela Ciaramidaro[3], Bruno G. Bara[1,4], Francesca M. Bosco[1,4]

Affiliations:

[1]University of Turin, Center for Cognitive Science, Dept. of Psychology, Turin, Italy

[2]Schulich School of Medicine and Dentistry, Dept. of Psychiatry, University of Western Ontario, London, Ontario, Canada

[3]Dept. of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Goethe-University, Frankfurt/M., Germany

[4] Neuroscience Institute of Turin, Italy

Corresponding author: Daniela Rabellino, Department of Psychiatry, London Health Sciences Centre, 339 Windermere Rd., London, Ontario, N6A 5A5, Canada. E-mail address: drabelli@uwo.ca

Abstract (165/300)

This study aims at exploring the decision-making process involved in third-party punishment within an economic frame, using the Third-Party Punishment game. We investigated altruistic punishment, i.e. the behavior of spending one's own money, with no personal gain, to punish those who violate the norms of cooperation. We analyzed this behavior in an in-group and out-group game setting, to compare how individuals behave with members of their own group (in-group) and with members of another group (out-group). In particular, our study proposed a comparison between two different cultures: Italian and Chinese. Our results confirmed the existence of altruistic punishment behavior in both experimental groups. In accordance with the parochial altruism theory, this tendency emerged as more prominent when faced with unfair play by the other group's members toward a member of one's own group. Furthermore, both groups exhibited a propensity for anti-social punishment behavior: many participants also spent small amounts of money to punish fair behavior, and this tendency was more evident in the Italian sample.

Introduction

The aim of the present study was to investigate cooperative and anti-social behavior that emerges through punishment in intra and inter-group settings. The novelty of the present study is to decline the intra and inter-group variable according to the real membership of the participants to their cultural group, i.e. Italians, example of individualist culture that emphasizes personal achievement and Chinese as example of collectivist culture, that emphasizes family and work group goals [1].

One of the most important features of human beings is the ability to cooperate, which allows us to build and to expand social groups up to a world-scale society. Cooperation is a crucial step for evolution, shaping the social norms which permit groups to survive and develop social institutions [2], [3]. Since adherence to social norms and rules plays a central role for the survival of social groups, offenders are punished for cheating [4]. As Fehr and Fischbacher (2004) [5] pointed out, norm enforcement occurs through sanctioning by individuals who punish violations of the expected behavioral standard. An interesting case is the Third-Party Punishment (TPP) game that involves a third player who observes an unfair interaction and spends part of his resources to maintain the rules, even though he is not directly affected by the violation. This form of behavior is defined as *altruistic punishment*: the costly punishment of norm violations, which does not involve any overt benefit for the punisher [6], [7], [8]. Strong reciprocity is the underlying force that sustains this behavior [9]. Strong reciprocators are prone to cooperate with other group members and to punish non-cooperators, even though this behavior is not sustained by any personal interest, kinship or reciprocal altruism. Altruistic punishment and strong reciprocity have been shown to play a positive role in maintaining cooperation, investigated through various economic game settings [10], [11], [12], [13], [14]. Cooperation and punishment can co-evolve when those who punish defectors are rewarded by being considered more trustful and worthy of cooperation from others [15], [16].

On the other hand, the literature has also reported the existence of an opposite and apparently paradoxical behavior, called anti-social punishment [17], [18]: the tendency of some people (when experimental conditions allow it) to spend their own money even to punish those who have been cooperative, a widespread tendency to punish pro-social cooperators. Rand et al. (2010) [19] analyzed the effect of anti-social punishment on the evolution of cooperation and showed that the inclusion of anti-social punishment destroys the

evolutionary success of cooperation. Anti-social punishment was excluded from previous models because of its incompatibility with the evolution of cooperation, but the fact that cooperators are in fact sometimes punished raises some interesting evolutionary questions. The possible explanations are far from conclusive, but some hypotheses have been formulated [20]: 1) retaliation, people who have been punished for being unfair in the past retaliate against cooperators, who are the people most likely to punish unfair behavior; 2) weak norms of civic cooperation and a weak rule of law, since strong norms of civic cooperation might act as a limit for anti-social punishment; 3) individual personality, such as strong taste for dominance or competitive personality or a desire to maximize relative payoffs of individuals; 4) measure of individualism-collectivism at a societal level [21]: people in collectivist societies might be more inclined to perceive other participants as out-group members. Hence, anti-social punishment might be stronger in collectivist societies than in individualist ones.

It follows that group affiliation may affect pro-social as well as anti-social behavior. In inter-group conflicts the *parochial social instinct* emerges, leading individuals to favor the members of their own ethnic, cultural or language group [22], [23]. Therefore in inter-group interactions people tend to protect or favor their own group members at the expense of those of other groups, even without any personal gain. There is a huge cross-societal variation in these complex behaviors [24], [25]. Following Hofstede's classification (1980) [26], different behaviors are exhibited by members of collectivist cultures that emphasize family and work group goals, whereas individualist cultures emphasize personal achievement (Ratner & Hui, 2003). Hence, we would expect collectivist groups to mete out stronger punishments to agents who break cooperative rules than individualist ones. Furthermore in-group favoritism is a spread behavior but a lot has still to be known about the magnitude of this tendency across societies. A study on in-group favoritism by Donna Harris et al. (2010) [28] showed that within a Thai subject pool emerged an "in-group bias" norm that is enforced within and

outside the group, while the Western sample (UK subject pool) did not show such a strong in-group favoritism. These results can lead our hypothesis to expect higher in-group favoritism within the Chinese sample compared to the Italian one. We investigated altruistic punishment and anti-social behavior by conducting a Third-Party Punishment experiment (TPP) within a Dictator Game [29], [30], [31]. In particular we analyzed the punishment behavior in observing different levels of fairness of the Dictator (Player A) according to the amount of money he transferred to Player B: fair, unfair and neutral. In-group and out-group settings were built on a cross-cultural context: Italian, example of an individualist culture, vs. Chinese, example of a collectivist culture [26]. This setting allowed us to observe behavior toward in-group versus out-group members. We used the term "in-group" to refer to the condition of interacting with the members of one's own cultural group, whereas the "out-group" condition refers to interaction with another cultural group's members.

This is the first study that investigates altruistic and antisocial behaviors through a TPP game where the facing groups are differentiated by their real cultural membership (Italian vs. Chinese).

According to the altruistic punishment theory [32], we hypothesized that members from both groups would punish unfair behavior. In line with parochial altruism however, we expected to find differences in the amount of money spent to punish, depending on in-group or out-group settings and, specifically, that higher amounts would be spent to punish members of the out-group. Moreover, we thought it likely that different behaviors would be affected by the specific cultures being investigated. We expected to observe a stronger tendency towards parochial altruism within the Chinese sample, belonging to a collectivist culture, than in the Italian group, belonging to an individualist culture. Italy is an example of an individualist, "me"-centered culture, especially in the biggest Northern cities where people can feel alone even in the middle of a big and busy crowd [21]. Collectivist cultures emphasize family and

work group goals while individualist cultures emphasize personal achievement [27], [33]. Hence we would expect collectivist groups to punish the violation of social norms more severely than individualist ones.

Finally, we wished to investigate whether participants belonging to a specific culture, i.e. Italian vs. Chinese, would punish anti-social behavior, and to what extent, even by spending some of their money to punish cooperative players.

Personality traits and characteristics i.e. honesty/humility and agreeableness [34], empathy skills [35], [36], and the tendency toward pro-social vs. competitive behavior [37], have also been found to be related to decision-making behaviors [38], [39], [40]. In order to rule out the possibility that individual differences have a crucial role in explaining our results, we took individual differences into account, investigating the personality traits of punishers, as well as their empathy skills and social orientation in relations with different cultures.

In summary, this study aimed to investigate altruistic punishment behavior using in-group and out-group conditions, with different levels of fairness exhibited by the dictator (fair, unfair or neutral). Therefore, the Third-Party Punishment (TPP) game within a cross-cultural context (Italian vs. Chinese) was used to test the following hypotheses:

1. both groups would show altruistic punishment behavior, punishing unfair behavior by the dictator (Player A);

2. different sanctions would be imposed (shown by the amount of money invested to punish) depending on in-group or out-group settings, i.e. higher amounts spent to punish members of the other group, pointing to the existence of parochial altruism;

3. participants belonging to a specific culture, i.e. Italian vs. Chinese, would punish and to what extent, even by spending some of their money to punish cooperative players;

4. finally, we evaluated the personal characteristics of the individuals participating in the present study in order to rule out the possibility that these could have played a crucial role in explaining our results.

Methods

Participants

Fifty-two male volunteers were recruited from among undergraduate students at the University and Polytechnic of Turin. The sample was composed of two groups: 26 subjects were belonging to the Chinese group and 26 subjects were belonging to the Italian group. Chinese students had been living in Italy for at least nine and not more than 24 months, (mean period of 15.9 months, SD = 5.8). Italian participants were 22.54 (±1.89) years old, whereas Chinese were 22.62 (±1.88) years, with no significant differences between the two samples (p=.899). None of the participants had previous experience of economic game experiments. The subjects received a lump sum payment of 20 euros for taking part in the experiment, plus the money they earned during the TPP game (from 0 to 20 euros). Inclusion criteria were a basic knowledge of English as verified by a test given immediately after the instructions, to measure their level of comprehension of the rules of the game and questions, since the instructions and the game were administered in the English language.

The study was approved by the Bio-Ethics Committee of the University of Turin. All participants gave their written informed consent following a detailed presentation of the study.

Procedure

We investigated TPP [41] behavior by adding a third player (C) to a classic Dictator Game, such as in Strobel et al. (2011) [42]. In a Dictator Game two players interact: player A, the dictator, and player B, the receiver. Player A has an initial endowment of points (20 points),

that he can share (or not) with Player B, whose role is passive. Player C observes the Dictator Game and he can use part or all of his endowment (4 points) to punish Player A's behavior. For every punishment point invested by Player C, 2.5 points are subtracted from Player A's payoff: for example if Player C investing all of his 4 points (the maximum punishment has been given) to sanction Player A's behavior, 10 points would be subtracted from Player A's payoff. We were interested in Player C's behavior, therefore our experimental subjects were always assigned to the role of the third party. The experiment began by placing all the partners in one room, for a total of five partners, but only one of the participants was a real player (the other four were all confederates of the experimenters). The experimental subjects were not aware that the other participants were confederates of the experimenter.

At the beginning, the experimenter read the general instructions and the five players were requested to listen to the instructions without speaking to each other. In order to create in-group and out-group settings, we brought together partners from both cultures, so the four confederates comprised two from the same cultural group as the subject and two from the other cultural group (i.e. two Italians and two Chinese). After been briefed, the five partners were conducted separately into individual rooms to play the game on a PC. Before starting, each subject read the instructions, written in English, concerning the role of the party he was playing and details about the game. The participants were then required to answer a questionnaire, which too was written in English, to assess their understanding of the rules. They were also trained by performing a short demonstration run to familiarize with the response buttons and paradigm.

The participants played the game, which was repeated many times (for a total of 120 trials). For each trial, Player C was always informed about the group affiliation of Players A and B. The subject (CHIN=Chinese or ITA=Italian) observed four different combinations of Players A and B depending on the in-group and out-group associations: A and B could be

both from Player C's group, or both from the out-group, or player A from the in-group and player B from the out-group, or vice-versa player A from the out-group and player B from the in-group.

The subjects believed they were playing with the other players via PC with Wi-Fi connection, although in actual fact they were faced with a programmed script (software E-Prime 2.0, 2007, Psychology Software Tools).

Material

We conducted 30 pseudo-randomized trials for each experimental in-group and out-group condition (for a total of 120 trials). Player A, with an initial endowment of 20 points, could behave fairly sending from 8 to 10 points to Player B (14 trials), unfairly sending from 0 to 6 points (15 trials) or in a neutral way sending 7 points to Player B (1 trial) (for fair-unfair cut-off see behavioral study in Strobel et al., 2011). Subsequently, Player C could decide for each trial whether or not to punish Player A's behavior.

Each game began by showing the group affiliation of Player A and Player B for three seconds. Then the subjects saw Player A's decision (2.5-5 seconds). They then had six seconds to punish A's behavior using a scale from 0 to 4 points (where for every punishment point they invested to punish Player A, 2.5 points were subtracted from Player A's endowment). Finally, the payoff for all Players, A, B and C, was revealed and displayed for three seconds. An additional fixation cross (four seconds) separated the outcome from the subsequent game (see Figure 1).

Insert Figure 1 about here

To check whether Player C had paid attention to the nationality of the other two players during the game, we asked the nationality of one of the two partners (A or B) playing in the previous trial (the question was asked six times in relation to Player A and six times in

relation to Player B). The subject knew that a wrong answer would result in resetting the points just earned in the last game to zero.

After the TPP session, we administered self-report questionnaires in the subjects' mother tongue. They completed the:

- Interpersonal Reactivity Index (IRI) to investigate patterns of social interaction [43], according to the following subscales: perspective-taking, fantasy, empathic concern, personal distress;
- Social Dominance Orientation (SDO) [44] to detect the basic desire for group-based, anti-egalitarianism and in-group dominance;
- Social Value Orientation (SVO) [45] to assess interpersonal orientation as pro-social, individualist, or competitor;
- HEXACO-PIr about personality traits [46], according to six scales: Honesty-Humility, Emotionality, eXtroversion, Agreeableness, Consciousness, and Openness to Experience.

Lastly, we administered the Implicit Association Test (IAT) on racial prejudice [47] to gauge prejudicial attitudes about Caucasian or Asiatic people. In the IAT the subject answers a series of items that have to be classified into four categories – typically, two representing a concept discrimination (Caucasian vs. Asiatic) and two representing an attribute discrimination such as *pleasant* versus *unpleasant*. The IAT produces measures derived from latencies of responses that are interpreted in terms of association strengths by assuming that subjects respond more rapidly when the concept and attribute mapped onto the same response are strongly associated than when they are weakly associated.

At the end of the experiment, the total points gained by Player C were converted into real money and paid to the subject. The subjects then left without meeting the other partners.

Data analysis

Altruistic Punishment Behavior

After an initial overview of the behavior of the whole sample in terms of frequency and average punishment behavior in the fair and unfair conditions, we conducted a repeated measure ANOVA with a 2 (players' C nationality: Italian vs. Chinese) x 4 (in-group and out-group settings: abC, ABC, AbC, aBC) x 3 (fair/unfair/neutral behavior by Player A, the dictator) design. The character case (upper case for in-group and lower case for out-group) was used to refer to the players A (the dictator) and B indicates the players' membership to a cultural group with respect to C: abC (A and B belonged to the out-group with respect to C); AbC (A belonged to the in-group and B to the out-group); ABC (A and B belonged to the in-group); aBC (A belonged the out-group and B to the in-group). The participants' nationalities were considered as a between-subjects factor, whereas in-group and out-group settings and fairness level were considered as within-subjects factors.

Parochial altruism

In order to specifically investigate altruistic punishment behavior we decided to isolate the unfair condition and differently grouped the data concerning in-group and out-group conditions, specifically separating Player A and Player B to better investigate Player C's with respect to A and B membership (as in-group or out-group members). Thus we conducted a separated repeated measures ANOVA design 2 (player C's nationality: Italian vs. Chinese) x 2 (player A's nationality: Italian vs. Chinese) x 2 (player B's nationality: Italian vs. Chinese) considering the unfair condition only.

Anti-social Punishment

To explore the anti-social punishment by Player C, the same repeated measures ANOVA was conducted as for the parochial altruism, but considering the fair condition only.

Bonferroni Post-hocs were calculated for significant effects and interactions in the previous ANOVAs. Last of all, descriptive analysis was conducted on psychological questionnaires, followed by T-tests to compare the two sample groups. Spearman's correlations were calculated between psychological data from questionnaires and separately with the fair condition and the unfair condition for each sample group.

Results

Altruistic Punishment Behavior

Our results showed that the majority of individuals in the experimental sample (96.15%) chose to punish the dictator (Player A) in the unfair condition. The mean number of points spent by Player C to punish in the unfair condition was 2.05 (±0.93) in the Italian sample, and 2.21 (±1.20) in the Chinese sample (Figure 2), with no significant difference between the two cultural groups (independent T-test: $t_{(50)}$=.524, p=.602, two-tailed).

To investigate the altruistic punishment behaviors the 3ways-ANOVA revealed a significant main effect of fairness ($F_{(2,100)}$=37.239, p<.001, partial η^2= .427) and emerged as the behavior of punishing Player A's unfairness more severely than Player A's fairness (Bonferroni post-hoc: p<.001) (Figure 2). There was no main effect of the in-group out-group settings variable ($F_{(3,150)}$=1.689, p=.178), as well as for Player C's nationality ($F_{(1,50)}$=1.076, p=.304). No interaction effects emerged for the interactions between in-group out-group settings and Player C's nationality, fairness level and Player C's nationality, fairness level and in-group out-group settings (all p≤271).

Insert Figure 2 about here

Parochial altruism

To conduct an in-depth analysis of Player C's tendency to punish Player A's unfair behavior in the in-group and out-group settings, we focused on the unfair condition - which favors punishment in respect of violation of social norms - and we conducted a repeated measures ANOVA with Player A's nationality (Italian vs. Chinese) and Player B's nationality (Italian vs. Chinese) as a within-subjects factors, and the participant's nationality as a between-subjects factor (Italian vs. Chinese). No main effects concerning Player A's nationality ($F_{(1,50)}$=1.050, p=.311) or Player B's nationality ($F_{(1,50)}$=.258, p=.614) emerged, but significant results emerged in the interaction between Player C's and Player B's nationalities ($F_{(1,50)}$=5.148, p=.028, partial η^2= .093) with a tendency to invest more points for punishment when Player C belonged to the same cultural group as Player B, as the victim of unfair behavior by an out-group Player A (T-test differences between the conditions aBC and abC paired t-test$_{(50)}$ =1.994, p=.052, two-tailed, and between aBC and AbC paired t-test$_{(50)}$=2.174, p=.034, two-tailed; Figure 3). By contrast, no main effect resulted on Player A's nationality ($F_{(1,50)}$=1.050, p=.311).

Insert Figure 3 about here

Anti-social Punishment

In order to analyze anti-social punishment, we focused on the fair condition, comparing the mean number of points spent by Player C to punish Player A, even if Player A had behaved fairly (by sending Player B from 8 to 10 of his points). An example of anti-social punishment was when Player A sent half of his points to Player B (fair behavior) and Player C decided to punish Player A.

96% of the Italian subjects resorted to anti-social punishment, with an average punishment score of 1.16 (±0.89) points. 76% of those in the Chinese group used anti-social punishment with a mean punishment score of 0.6 (±0.81) points. A repeated measures ANOVA was conducted to investigate the association between the nationalities of Player A's nationality (Italian vs. Chinese) and Player B's nationality (Italian vs. Chinese) as a within-subjects factor, and the study subject's nationality as a between-subjects factor (Italian vs. Chinese).

No significant main effects of the nationalities of Player A ($F_{(1,50)}$=.047, p=.830) and Player B ($F_{(1,50)}$=.013, p=.911) emerged, nor did any significant interactions between the nationalities of Player A, Player B and that of the study subject (all p≥.123). However, a main effect of Player C's nationality as between-subjects factor was significant ($F_{(1,50)}$=48.675, p=.048, partial η^2=.076) since the Italian sample used consistently more points than the Chinese one did for punish fairly behavior (Bonferroni post-hoc on Player C's nationality Italian vs. Chinese p=.048) (see Figure 4).

Insert Figure 4 about here

Psychological traits

We also examined whether psychological features correlated with altruistic punishment behavior in Player C. Correlations were calculated between psychological questionnaires and separately with the fair condition and the unfair condition. The outcomes for both the Chinese and the Italian sample on all administered questionnaires are shown in the following table1.

Insert table1 about here

Independent T-tests between the Chinese and the Italian groups revealed significant differences on the Agreeableness subscale for personality characteristics evaluated using the HEXACO (Chinese vs. Italians: t=3.280, p=.002), on the personal distress subscale of the IRI questionnaire, evaluating the tendency to experience distress or discomfort in response to extreme distress in others (Chinese vs. Italians: t=3.792, p<.001) and on Social Dominance Orientation (Chinese vs. Italians: t=3.714, p=.001).

Regarding preference for one's own racial group (Caucasian vs. Asian) measured with the IAT, both groups preferred their own racial group compared to the other racial group (for the interpretation of the significance of *d,* see Cohen, 1988) [48], but Italians showed a significantly higher preference for their own racial group (Chinese vs. Italians: t=-2.646, p=.011) (TAB.2).

Insert Table 2 about here

Analysis of the SVO questionnaire on the subjects' interpersonal orientation toward others placed 34.5% of Italians in the individualist category, 46% in the pro-social category and left 19.5% uncategorized. Within the group of Chinese subjects, 30.8% were placed in the individualist category, 3.8% in the competitive category, 42.3% were categorized as pro-social and the remaining 23.1% were uncategorized. No significant differences resulted from a comparison between the two sample groups (Pearson χ^2=1.193, p=.755).

We then examined the correlation between punishment behaviors and the two sample groups. In the Chinese group there was only a significant correlation between punishment in

the unfair condition and one of the HEXACO subscales (HEXACO_Emotionality) (Spearman's R=-.409, p=.038). None of the other subscales correlated with punishment scores (all p≥.201); none of the scores on the IRI, SVO, SDO or IAT revealed any significant results correlated to punishment behavior within the Chinese sample (all p≥.086).

Within the Italian sample, our findings showed a significant correlation between HEXACO_Emotionality and punishment in the fair condition (Spearman's R=.62, p=.001). On the other HEXACO subscales we only observed marginally significant correlations between the mean score for punishment in the unfair condition and HEXACO-Honesty/Humility (Spearman's R=.369, p=.063) and HEXACO-Openness to Experience (Spearman's R=.386, p=.052). The other HEXACO subscales showed no significant correlations (all p≥.277). On the IRI scales, the more subjects were used to experiencing personal distress (feelings of fear, apprehension and discomfort at witnessing the negative experiences of others) the more anti-social behavior they demonstrated, punishing Player A though fair (Spearman's R=.516, p=.007). The remaining questionnaires (SVO, SDO and IAT) showed no significant results related to punishment behavior within the Italian sample (all p≥.128).

Discussion

The aim of the present study was to investigate cooperative and anti-social behavior during the TPP economic game within a context of in-group (i.e. all participants belong to the same cultural group) and out-group settings (i.e. one of the participants belonged to a different cultural group), and to explore these complex behaviors comparing groups from different cultural backgrounds, and specifically Italians vs. Chinese.

Our first hypothesis concerned the propensity of a third party to spend own resources punishing an observed unfair behavior, even when not directly affected by that behavior. This

study shows that, in both in-group and out-group settings, the third party acts as a strong reciprocator, choosing to punish unfair behavior by the dictator (Player A) even if that means using his own resources, in line with the theory of altruistic punishment [9], [32], [12]. Members from both groups reacted to the unfair condition in the same way, with no significant differences being observed between both. This result supports the idea that altruistic punishment is a cross-cultural tendency in larger societies [49], [50].

For our second hypothesis, we expected that parochial altruism would emerge as the tendency to protect one's own group members by delivering harsh punishment, and thus that the punishing behavior of Player C would be influenced by the in-group or out-group setting [51], [52]. Our analysis produced significant results on this issue, since both Chinese and Italian participants invested more points to punish unfair behavior by the dictator when the victim was an in-group member. The condition of observing an in-group member as the victim of unfair behavior seemed to make the difference in inter-group settings, whereas Player A's (the dictator's) membership did not influence Player C's punishment behavior. These results account for cooperative behavior, since a judge observing unfair behavior is inclined to sacrifice his own resources to seek revenge for the victim and protect him from further damage. Adherence to and enforcement of social norms can explain such altruistic behavior aimed toward construction and conservation of civic cooperation [20]. Such behavior appears to be in line with a tendency for in-group favoritism [53].

For what concerns altruistic behavior no differences concerning the membership to Italian vs. Chinese culture emerged. Different explanations can relate to these results: Chinese and Italian cultures can be not so far along collectivism/individualism continuum line. The development of Chinese culture toward consumeristic lifestyle, especially in big cities, can have reduced the gap, and Hofstede's classification is far from conclusive and completely accepted in recent debates [54], [55]. Moreover the fact that Chinese participants were

actually immigrants embedded within Italian culture can have narrowed the differences and be biased by social desirability, intended as the tendency of giving a socially accepted representation of themselves [56]. Thus we have to consider the results in the unfair condition as influenced by ingroup bias, but not by cultural membership.

Finally, we were also interested, to investigate the anti-social punishment behavior. To analyze anti-social punishment behavior we had to focus on the fair condition, i.e. when Player A, the dictator, transferred to B from 8 to 10 points. We observed that both cultural groups exhibited anti-social punishment behavior, spending their own resources to punish cooperators. These results remind to the literature that refers to antisocial punishment as a puzzling behavior [19], indeed the same participants showing altruistic behavior, switch to an antisocial behavior when conditions allow it.

Furthermore interestingly, such punishment was significantly harsher in the Italian sample than in the Chinese one. Excluding retaliation, since this condition was not possible in our experimental design, there are some possible explanations for this behavior. According to Herrmann (2008) [20], strong norms of civic cooperation, such as the strength of the rule of law, might restrain anti-social punishment. We would therefore expect to find higher levels of anti-social punishment among participants from societies with weak norms of civic cooperation and a weak rule of law. From that perspective, the results of this study would lead to consider Italian norms of civic cooperation weaker than Chinese ones. Since punishment in our experiment was cheaper for the punisher than for the dictator, another possible explanation for anti-social punishment is that a competitive personality [57], or taste for dominance [58] might drive individuals to punish pro-social cooperators. However Italians, who used anti-social punishment more than Chinese, did not show higher score on psychological questionnaires in relation to these traits.

Finally for what concerns antisocial punishment we did no found differences in in-group and out-group conditions. Such result is quite surprising at least for Italian group since we found that the Italian sample (with respect the Chinese one) showed a significant higher preference for their own racial group (IAT evaluation). Our findings on antisocial punishment are far to be conclusive and further research is necessary in order to clarify such puzzling behavior.

For what concerns the remaining personal characteristics of the individuals participating in the present study we found that participants in the Chinese group scored significantly higher on the Agreeableness subscale of the HEXACO questionnaire and on the personal distress subscale of the IRI questionnaire, leading us to argue that our Chinese sample was more empathic and more agreeable than the Italian one. Moreover, the Chinese group exhibited a greater tendency to consider their social group as predominant (SDO questionnaire). The SVO and other subscales of the HEXACO and IRI did not reveal any other differences between the two groups. Although we found some significant differences between the two sample groups in terms of psychological characteristics, none of these is able to account for the punishment behaviors observed in the two cultural groups in this study.

Concerning the correlation between psychological characteristics and punishment behavior, only four significant results emerged: HEXACO_Emotionality negatively correlated with altruistic punishment in the Chinese group and positively correlated with anti-social punishment in the Italian group; IRI_Personal distress and HEXACO_Openness to experience positively correlated with anti-social punishment in the Italian sample; no other correlations reached significance. Globally considered we cannot consider personality characteristics and racial bias as possible explanations for our results on altruistic and anti-social punishment behaviors in in-group and out-group settings.

A limit of the present study is the small sample size and it would need to be enlarged to enable any generalization of our results, and a real cross-cultural setting would better explain the differences between collectivist and individualist societies (even the higher scores obtained by the Chinese sample on the agreeableness scale could be explained by the desire of immigrants to appear socially accepted). However, bearing these limitations in mind, our findings support those of previous studies on altruistic and anti-social punishment [17], [19], [59], [60], [61].

Moreover, this is the first time that punishment by third parties has been studied in real groups formed on a cultural basis and using an experimental design that permits both altruism and anti-social behavior. Our study raises new questions on group dynamics and in-group and out-group settings: Is the condition of an in-group member as a victim more emotionally activating than that of a dictator? What are the distinctive features of an individualist/collectivist culture? What are the implications of in-group bias on anti-social behavior? A possible challenge for future researches will be to answer to these questions.

Further empirical and theoretical investigation of in-group favoritism could drive future research on the evolution of cooperation and anti-social behavior.

Acknowledgments

The authors acknowledge work of undergraduate students for their technical assistance, as well as the confederates of the experimenters for their participation.

References

1. Ratner C, Hui L (2003) Theoretical and Methodological Problems in Cross-Cultural Psychology. J Theory Soc Behav 33(1): 67-94. doi: 10.1111/1468-5914.00206.
2. Tomasello M (2009) Why we cooperate. Cambridge: The MIT Press. 208 p.

3. Helbing D, Szolnoki A, Perc M, Szabo´ G. (2010) Evolutionary establishment of moral and double moral standards through spatial interactions. PLoS Comput Biol 6(4), e1000758. doi: 10.1371/journal.pcbi.1000758.

4. Fehr E, Fischbacher U, Gächter S (2002) Strong reciprocity, human cooperation, and the enforcement of social norms. Human Nature 13(1): 1-25.

5. Fehr E, Fischbacher U (2004) Third-party punishment and social norms. Evol Hum Behav 25(2): 63-87.

6. Fehr E, Gächter S (2000) Fairness and retaliation: The economics of reciprocity. The journal of economic perspectives 14(3): 159-181. doi: 10.1257/jep.14.3.159.

7. Gintis H (2000) Strong reciprocity and human sociality. J Theor Biol, 206(2): 169-179.

8. Jensen K (2010) Punishment and spite, the dark side of cooperation. Philos Trans R Soc Lond B Biol Sci 365(1553): 2635-2650. doi: 10.1098/rstb.2010.0146.

9. Bowles S, Gintis H (2004) The evolution of strong reciprocity: cooperation in heterogeneous populations. Theor Popul Biol 65(1): 17-28.

10. Gintis H, Bowles S, Boyd R, Fehr E (2003) Explaining altruistic behavior in humans. Evol Hum Behav 24(3): 153-172.

11. Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, McElreath R (2001) In search of homo economicus: behavioral experiments in 15 small-scale societies. Am Econ Rev 91(2): 73-78.

12. Fehr E, Fischbacher U (2004) Social norms and human cooperation. Trends Cogn Sc 8(4): 185-190.

13. Boyd R, Gintis H, Bowles S (2010) Coordinated punishment of defectors sustains cooperation and can proliferate when rare. Science 328(5978): 617-620. doi: 10.1126/science.1183665.

14. Shinada M, Yamagishi T, Ohmura Y (2004) False friends are worse than bitter enemies:“Altruistic” punishment of in-group members. Evol Hum Behav 25(6): 379-393.

15. Gardner A, West SA (2004) Cooperation and punishment, especially in humans. Am Nat 164(6): 753-764.

16. Krasnow MM, Cosmides L, Pedersen EJ, Tooby J (2012) What are punishment and reputation for? PLoS One 7(9):e45662. doi: 10.1371/journal.pone.0045662.

17. Rand DG, Nowak MA (2011) The evolution of antisocial punishment in optional public goods games. Nat Commun, 2: 434. doi: 10.1038/ncomms1442.

18. Dreber A, Rand DG (2012) Retaliation and antisocial punishment are overlooked in many theoretical models as well as behavioral experiments. Behav Brain Sci 35 (1): 24. doi: 10.1017/S0140525X11001221.

19. Rand DG, Armao IV J J, Nakamaru M, Ohtsuki H (2010) Anti-social punishment can prevent the co-evolution of punishment and cooperation. Journal of theoretical biology 265(4): 624-632. doi: 10.1016/j.jtbi.2010.06.010.

20. Herrmann B, Thöni C, Gächter S (2008) Antisocial punishment across societies. Science 319(5868): 1362-1367. doi: 10.1126/science.1153808.

21. Hofstede G (2001) Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. Thousand Oaks (CA): Sage CA. 616p.

22. Bernhard H, Fischbacher U, Fehr E (2006) Parochial altruism in humans. Nature 442(7105): 912-915.

23. Sääksvuori L, Mappes T, Puurtinen M (2011) Costly punishment prevails in intergroup conflict. Proc Biol Sci 278(1723): 3428-3436. doi: 10.1098/rspb.2011.0252.

24. Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, et al. (2005) "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. Behav Brain Sci 28(6): 795-814.

25. Henrich J, McElreath R, Barr A, Ensminger J, Barrett C, Bolyanatz A, et al. (2006) Costly punishment across human societies. Science 312(5781): 1767-1770.

26. Hofstede G (1980) Culture's consequences: international differences in work-related values. Newbury Park, CA: Sage. 328p.

27. Ratner C, Hui L (2003) Theoretical and Methodological Problems in Cross-Cultural Psychology. J Theory Soc Behav 33(1): 67-94.

28. Harris D, Herrmann B, Kontoleon A (2010) The Impact of Costly Punishment on In-group Favouritism Experimental Evidence from Thailand and the UK. Working paper available on http://www.economics.ox.ac.uk/index.php/academic/donna-harris

29. Almås I, Cappelen AW, Sørensen EØ, Tungodden B (2010) Fairness and the development of inequality acceptance. Science 328(5982): 1176-1178. doi: 10.1126/science.1187300.

30. Fehr E, Falk A (2002) Psychological foundations of incentives. Eur Econ Rev 46: 687.

31. Eckel C, Grossman P (1996) Altruism in anonymous dictator games. Games and Economic Behavior 16: 181.

32. Fehr E, Gächter, S (2002) Altruistic punishment in humans. Nature 415(6868): 137-140.

33. Triandis HC, Brislin R, Hui CH (1988) Cross-cultural training across the individualism-collectivism divide. Int J Intercult Relat 12(3): 269-289.

34. Hilbig BE, Zettler I (2009) Pillars of cooperation: Honesty–Humility, social value orientations, and economic behavior. J Res Pers 43(3): 516-519.

35. Brüne M, Scheele D, Heinisch C, Tas C, Wischniewski J, Güntürkün O (2012) Empathy Moderates the Effect of Repetitive Transcranial Magnetic Stimulation of the

Right Dorsolateral Prefrontal Cortex on Costly Punishment. PloS one 7(9), e44747. doi: 10.1371/journal.pone.0044747

36. Singer T, Seymour B, O'Doherty J P, Stephan KE, Dolan RJ, Frith C D (2006) Empathic neural responses are modulated by the perceived fairness of others. Nature 439(7075): 466-469.

37. Dana J, Cain D M, Dawes R M (2006) What you don't know won't hurt me: Costly (but quiet) exit in dictator games. Organ Behav Hum Decis Process 100: 193–201.

38. Ashton D (1998) Systematic risk and empirical research. Journal of Business Finance and Accounting 25 (9/10): 1325–1356.

39. Boone C, De Brabander B, Van Witteloostuijn A (1999) The impact of personality on behavior in five Prisoner's Dilemma games. J Econ Psychol 20(3): 343-377.

40. Ben-Ner A, Kramer A (2011) Personality and altruism in the dictator game: Relationship to giving to kin, collaborators, competitors, and neutrals. Pers Individ Dif 51(3): 216-221.

41. Greif A (1994) Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies. J Polit Econ 102: 912–950.

42. Strobel A, Zimmermann J, Schmitz A, Reuter M, Lis S, Windmann S, Kirsch P (2011) Beyond revenge: Neural and genetic bases of altruistic punishment. Neuroimage 54(1): 671-680. doi: 10.1016/j.neuroimage.2010.07.051.

43. Davis MH (1983) Measuring individual differences in empathy: Evidence for a multidimensional approach. J Pers Soc Psychol 44(1): 113-126.

44. Sidanius J, Pratto F (1999) Social dominance: An intergroup theory of social hierarchy and oppression. New York: Cambridge University Press. doi: 10.1002/9780470672532.wbepp253.

45. Van Lange PAM (1999) The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. J Pers Soc Psychol 77: 337-349.

46. Ashton MC, Lee K (2009) The HEXACO-60: A short measure of the major dimensions of personality. J Pers Assess 91:340-345.

47. Greenwald AG, McGhee DE, Schwartz JLK (1998) Measuring individual differences in implicit cognition: The Implicit Association Test. J Pers Soc Psychol 74:1464-1480.

48. Coehn J (1988) Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum. 567p.

49. Marlowe FW, Berbesque JC, Barr A, Barrett C, Bolyanatz A, Cardenas JC, et al. (2008) More 'altruistic'punishment in larger societies. Proc Biol Sci 275(1634):587-590.

50. Marlowe FW, Berbesque JC, Barrett C, Bolyanatz A, Gurven M, Tracer D (2011) The 'spiteful'origins of human cooperation. Proc Biol Sci 278(1715): 2159-2164.

51. Brewer MB (1979) In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. Psychol Bull 86(2):307-324.

52. Baumgartner T, Götte L, Gügler R, Fehr E (2012) The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. Hum Brain Mapping 33(6): 1452-69. doi: 10.1002/hbm.21298.

53. Tajfel H, Turner JC (1986) The social identity theory of intergroup behavior. In Worchel S, Austin W, editors. Psychology of intergroup relation. Chicago: Nelson Hall. Pp. 7–24.

54. Voronov M, Singer JA (2002) The myth of individualism-collectivism: A critical review. J Soc Psychol 142(4):461-480.

55. Hamamura T (2012) Are cultures becoming individualistic? A cross-temporal comparison of individualism–collectivism in the United States and Japan. Pers Soc Psychol Rev 16(1):3-24.

56. Crowne D P, Marlowe D (1960) A new scale of social desirability independent of psychopathology. J Consult Psychol 24(4):349.

57. Clutton-Brock TH, Parker GA (1995) Punishment in animal societies. Nature 373(6511):209-216.

58. Liebrand WB, Jansen RW, Rijken VM, Suhre CJ (1986) Might over morality: Social values and the perception of other players in experimental games. J Exp Soc Psychol 22(3):203-215.

59. Gächter S, Herrmann B (2009) Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. Philos Trans R Soc Lond B Biol Sci 364(1518):791-806. doi: 10.1098/rstb.2008.0275.

60. Rand DG, Dreber A, Ellingsen T, Fudenberg D, Nowak, MA (2009) Positive interactions promote public cooperation. Science 325(5945):1272-1275. doi: 10.1126/science.1177418.

61. Egas M, Riedl A (2008) The economics of altruistic punishment and the maintenance of cooperation. Proc Biol Sci 275(1637):871-878. doi: 10.1098/rspb.2007.1558.

Figure legends

Figure 1. **Third Party Punishment Game**. Overview of one trial run of the TPP game used in the present study.

Figure 2. **Altruistic Punishment (Player C's mean punishment score).** In both cultural groups Player C spent higher amounts to punish in the unfair condition compared to the fair and neutral conditions. Error bars: standard error.

Figure 3. Unfair condition. Player C's punishment means score when Player B's nationality is Chinese or Italian. Error bars: standard error.

Figure 4: Fair Condition. Player C's mean punishment score when both A and B belonged to the in-group (ABC), when A belonged to the in-group and B to the out-group (AbC), when both A and B belonged to the out-group (abC) and when A belonged the out-group and B to the in-group (aBC). Error bars: standard error.

Tables

TAB.1 Psychological variables scores.

Questionnaires	CHIN (M±SD)	ITA (M±SD)	Independent T-Test
HEXACO_Honesty/Humility	3.29±.63	3.57±.78	t =-1.442; p =.156
HEXACO_Empathy	3.07±.45	2.88±.66	t = 1.178; p =.244
HEXACO_eXtroversion	3.57±.54	3.61±.44	t = -.282; p =.779
HEXACO_Agreeableness	3.53±.44	3.11±.49	t = 3.280; p =.002**
HEXACO_Conscientiousness	3.07±.48	3.39±.54	t = -.245; p =.807
HEXACO_Openness to experience	3.36±.34	3.54±.59	t = .365; p =.717
IRI_personal distress	3.05±.57	2.45±.55	t = 3.792; p<.001**
IRI_empathic concern	3.76±.51	3.58±.49	t = 1.299; p =.200
IRI_fantasy	3.07±.52	3.49±.65	t = .833; p =.409
IRI_perspective-taking	3.64±.59	3.19±.56	t = 1.320; p =.193

SDO	53.5±5.77	40.86±16.37	t = 3.714; p=.001**

Means ± SD of the HEXACO subscales, of the IRI subscales and of SDO, divided by group affiliation. The third column shows independent T-Test results and the significance level of the comparison between the Italian and the Chinese samples.

TAB.2: Racial prejudice.

	CHIN (Cohen's *d* ± SD)	**ITA (Cohen's *d* ± SD)**	**Independent T-Test**
IAT	*d*=.17±.47	*d*=.49±.35	t = -2.646; p =.011**

Coefficient values of IAT scores, followed by standard deviation for each sample. The third column shows independent T-Test results and the significance level of the comparison between the Italian and the Chinese samples.